

QÜESTIÓ, vol. 22, 2, p. 365-378, 1998

## ANALYSIS OF SURVEY DATA INVESTIGATING THE MALARIAL ENDEMICITY OF A MIXED TRIBAL POPULATION OF BIHAR, INDIA

T.K. BASU

S. GANGULY

S.K. SARKAR

Indian Statistical Institute\*

R. ARNAB

University of Durban\*\*

*A section of the mixed tribal population of the Singhbhum district, Bihar, India is declared malaria epidemic zone. The tribal population of several generation is known to be suffering from malaria. A survey based on cross-sectional data analysis, was conducted on the mixed tribal population for one month. The purpose of this study was to investigate the health status using collected blood samples. The main focus is on the comparative roles of the «defense mechanism» and «vitality factor» of the human system in context to the malarial infection. By gradual elimination of the blood parameters by statistical analyses, the «vitality» parameters probing malarial endemicity are assessed with a view to predicting the epidemic.*

*The main findings from this survey are (i) of the selected twenty two parameters of blood, albumin, total cholesterol, total protein,  $\beta$ -Globulin,  $\gamma$ -Globulin, Immuno globulin G seem to have some predictive capacity with respect to the malarial endemicity of the tribal people and (ii) Categorical variables like blood groups and sex are comparatively less important for prediction of malaria.*

**Keywords:** Vitality factors; defense mechanism; two-way Anova; discriminant function.

---

\* Biometry Research Unit. Indian Statistical Institute. 203 Barrackpore Trunk Road. Calcutta 700 035, India.

\*\* Department of Statistics. University of Durban, South Africa.

– Received January 1997.

– Accepted November de 1997.

## 1. INTRODUCTION

The Indian Council of Medical Research, a leading medical research concern of India, identified a few villages of tribal habitat of the Singhbhum district of Bihar (India) as a malarial epidemic zone. The people of these villages show evidence of malarial suffering for several generations. A study was undertaken to investigate the malarial endemicity by checking the blood samples obtained from a mixed tribal population of the Singhbhum district. The investigation was executed by the Indian Council of Medical Research in conjunction with the Biometric Unit, Indian Statistical Institute.

Malaria depends upon two principal factors- (a) vitality and (b) defense mechanism, regulated by the circulation of blood. Vitality is controlled by total, free and ester cholesterol, albumin, red blood corpuscles (RBC) and haemoglobin, and the defense mechanism is controlled by immunoglobulins (IgG, IgA and IgM), white blood corpuscles (WBC), monocytes, neutrophils, eosinophils and lymphocytes. Lymphocytes are subdivided into T and B lymphocytes. In addition, the role of globulin-containing fractions is also prominent on this aspect. Under malarial conditions, RBC, haemoglobin and albumin decrease whereas WBC, neutrophils and globulins tend to increase. This is a very broad overview of this complex mechanism. In this context, a brief description of the biology of malaria is presented.

### 1.1. Biology of Malaria

Malaria is caused by a parasite protozoan infection. Four different species of the genus *Plasmodium* are known to infect humans. In the tropics, *Plasmodium falciparum* is most prevalent. The life cycle of the parasite is an interaction between a female mosquito of the genus *Anopheles* (the vector) and the human host. Transmission of disease occurs by a bite of the infectious mosquito. The parasite migrates to the liver, remains in the latent stage for several days while replicating. Ultimately, there occurs a penetration of host RBC with an asexual replication within the parasites which results in the lysis of the cells. Malarial symptoms are occurred by this asexual parasites in the blood. When there is a fresh case of mosquito biting, a sexual stage, called *gametocytes* develops (from the asexual parasites) which is responsible for transmission of parasites from the host. These transmitted parasites fertilize in mosquito gut, replicate and a new cycle of transmission starts. These types of epidemiological surveys are generally based on blood smears in which one can observe both the *asexual* parasites and *gametocytes*. Classifications of blood smears are generally done in terms of the appearance/absence of parasite or in terms of its density. According to above, there are actually three classes of population related to the malarial status- susceptibles, the proportion that are uninfected (Class 1); infecteds, the proportion with infection (Class 2); and immunes, the proportion with asymptomatic infection which is not prominent

(Class 4). The Febrile Class i. e. Class 3 represented subjects with unidentified fever, not malaria (as confirmed by slide test).

## **1.2. Main Findings**

Subsequent analysis of data gives some indication that the levels of albumin, total cholesterol, total protein,  $\beta$  - Globulin,  $\gamma$  -Globulin and IgG in the circulating system may be used for the prediction of malarial endemicity of the tribal population. The categorical variables like blood groups and sex are less important for the prediction of the malarial incidence.

## **2. SAMPLING DESIGN AND COLLECTION OF DATA**

A simple random sample of 88 persons were selected from the list of volunteers from the tribal population. Among them, 45 were male and 40 were female. The selected individuals were invited at the Clinics of Kokda and Khanderber, organized by the Gram Vikas Kendra, a rural development centre of Tata Engineering and Locomotive Works, Jamshedpur, Bihar. The list of volunteers was prepared earlier by the Centre. 10 persons were invited daily in the Clinics serially from the chart. The average daily attendance was six. This is because most of the females did not respond owing to low literacy level, social-taboo or customs. The nature of nonresponse is quite random in nature and so this will not cause any bias in the experiment. However, the efficiency of the estimators will reduce because of the reduction of sample size. Relevant history regarding family members, total income of the family and other socio- economic data of the volunteers were collected by the Centre and were sent to the Clinics for record. They were mostly nonvegetarians residing in huts as usually observed in tribal villages. According to their family income, they belonged to the «below poverty-level» group.

### **2.1. Experimental protocol**

Subjects, as instructed earlier, came in post-absorptive stage (overnight fasting) in the Clinics between 9 to 10 am. The attending physician checked the weight, pulse and blood pressure of each individual and drew blood from their bracial vein for subsequent testing. 10ml of blood was taken in an aseptic condition for the biochemical, haematological and immunological tests. Out of this 10ml blood, 4ml was kept in a sterilized glass vial with the sequestering agent. Blood smear was prepared on two glass slides, one thick (for the detection of Malarial parasite) and the other thin

(for differential count of the blood cells). The remaining portion of blood was taken in a sterilized test tube and was allowed to clot for release of serum.

The electrophoretic separation of different protein fractions was carried out in the line of Smithies(1955) with 1% agar and 0.1M Veronal Buffer (pH 8.6). The fractions of Protein- Albumin and Globulins were scanned by using a Densitometer. The relative percentage of different Protein fractions was measured by using a Planimetre.

The Immunodiffusion study was performed according to the standard method (Mancini et al, 1965). The diameter of the precipitating ring was measured in mm by using the 'Immunomeasure' scale. The standard curve was plotted on a double log graph paper using standard antigens supplied by M/s Immunodiagnosics Pvt. Ltd., Delhi. The Immunoglobulin fractions were calculated from the standard curve. The conventional methods for the estimation of Total protein (Lowry et al, 1951), Total, Free and Ester cholesterol(Wootton,1974), Haemoglobin, RBC,WBC,Blood groups and Differential counts i.e.estimation of Leucocytes, Monocytes, Eosinophils, Basophils (Davie and Lewis,1984), T- and B- Lymphocytes (Blood, 1977) were followed (Summary tables 1A-1C). On the basis of the above pathological tests the population was post stratified into four Classes according to their malarial status.

Class 1. Normal: Those who had no fever currently and no past history of malaria within the last couple of years. Total number 20 (17 male and 3 female)

Class 2. Malarious: Those who turned up with symptoms of malaria and showed the malarial parasite positive in the 'slide test'. Total number 15 (11 male and 4 female)

Class 3. Febrile: Those who emerged with fever but no malarial parasite was identified in the 'slide test'. Total number 25 (15 male and 4 female).

Class 4. Dormant: Those who had recurring malaria during the last two years but showed no current malarial symptoms. Total number 28 (21 male and 7 female).

### **3. STATISTICAL ANALYSIS**

#### **A: Analyses of cardiovascular and blood parameters**

The means of the following twenty one blood parameters (viz. Total Protein (TP), Total Cholesterol (Tch), Free Cholesterol (Fc), Ester Cholesterol (Ec), Red Blood Corpuscles (RBC), White Blood Corpuscles (WBC), Neutrophils (N), Lymphocytes (L), Eosinophils (E), Monocytes (M), Haemoglobin (Hb), Immuno- globulin G (IgG), Immunoglobulin A (IgA), Immunoglobulin M (IgM), Albumin (Alb),  $\alpha_1$ -Globulin ( $\alpha_1$ ),  $\alpha_2$ -Globulin ( $\alpha_2$ ),  $\beta$ -Globulin,  $\gamma$ -Globulin, T-cell (Tc), B-cell (Bc) and two cardio-

vascular parameters, namely, Pulse (P) and Blood Pressure, Systolic/Diastolic [Bp (Sys/Dia)] for the four classes of people (Normal, Malarious, Febrin and Dormant) were estimated through the sample means which are given in the tables 1A, 1B and 1C. The estimated standard errors (SR) of the estimates are also presented in these tables. The following table gives an idea of the general health condition of the tribal community under consideration.

**Table 1**

Parameters	Normal Range Indian Standard	mean	standard deviation	standard error(se)	Proportion below normal range	Proportion above normal range	Proportion outside normal range	sample range min-max
P(c/min)	65-85	94.898	15.834	1.688	0	70.11	70.11	68-124
BP/sys	100-140	110.932	13.118	1.398	17.04	0	17.04	78-150
BP/dia	70-90	76.33	10.783	1.149	20.45	4.54	24.94	54-100
WBC(m/cmm)	4000-10000	7185.8	931.853	99.336	0	0	0	5000-9500
$\alpha_1$ (% Tp)	52-68	3.555	1.351	0.144	26.13	23.86	50	0.9-8.7
$\alpha_2$ (%Tp)	6.1-10.1	3.838	1.697	0.181	89.7	0	89.7	1-9.9
$\beta$ (% Tp)	8.5-14.5	6.619	2.495	0.266	84.1	0	84.1	2.4-14.5
$\gamma$ (% Tp)	10-21	18.991	5.38	0.537	4.54	27.27	31.88	6-42.9
IgG(mg%)	700-1500	2453.69	598.515	63.802	0	85.22	85.22	1205-3477
IgA(mg%)	90-450	193.818	18.698	1.993	0	0	0	152-237
IgM(mg%)	40-250	128.352	28.642	3.053	0	0	0	69-192
Alb(% Tp)	52-68	67.436	7.046	0.751	2.27	51.13	53.46	45.7-85.1
N(%)	60-65	56.125	5.858	0.624	65.9	6.81	72.72	42-68
L(%)	20-40	33.318	4.797	0.511	0	7.95	7.95	24-48
E(%)	1-3	8.148	4.268	0.455	0	87.45	87.45	1-19
M(%)	2-8	2.614	1.309	0.14	13.63	0	13.63	1-6
Hb(gm%)	14-16	11.063	1.231	0.131	100	0	100	8-12.8
Tc(%)	70-75	63.5	6.447	0.687	80.68	0	80.68	42-75
Bc(%)	15-20	24.08	5.126	0.546	1.13	70.45	71.59	12-38
Tp(gm%)	6-8	7.142	0.387	0.041	0	0	0	6.5-8
Tch(mg%)	150-280	126.773	22.951	2.447	82.95	0	82.95	80-184
Fc(mg%)	50-70	42.045	13.597	1.449	78.4	2.27	80.68	20-94
Ec(mg%)	95-210	85.886	22.21	2.368	69.31	0	69.31	26-140
RBC(m/cmm)	4000-10000	3.794	0.474	0.051	98.86	0	98.86	2.31-4.78

From the above table, we briefly comment on the average health condition as follows:

BP: Due to the simplicity of living and diet, the community under study maintained acceptably good BP levels.

RBC, Hb: Lower values indicate that most of the people are anaemic.

Tp, Tch, Fc, Ec: Lower values than the normal range is indication of low fat diets coupled with normal protein levels.

E: High concentration of eosinophils is generally correlated with parasitic infection in the people.

WBC, N, L, M: The levels of these parameters are acceptable in terms of facilitating immuno-defence mechanism.

**Summary Table: 1A**

Class	Sample size	Statistic	P (count/min)	BP/sys (mm Hg)	BP/dia (mm Hg)	TP (gm%)	Tch (mg%)	Fc (mg%)	Ec (mg%)	RBC (m/cumm)	WBC (no/count)
1	20	mean	91.6	112.6	77.7	7.08	136.2	39.2	102.5	4.062	7707.5
		sd	16.28005	8.9241	8.97273	0.32031	21.7154	11.26765	14.42047	0.28971	816.13035
		se	3.64033	1.9955	2.00636	0.07162	4.85572	2.51952	3.22451	0.06478	182.19229
2	15	mean	94.06667	108.67	7606	6.94666	96	38.63333	57.46667	3.72133	6736.66667
		sd	15.48533	13.656	12.04326	0.23907	10.0133	14.50455	13.9421	0.48780	772.97405
		se	3.99829	3.526	3.10955	0.06172	2.58543	3.74506	3.59983	0.12595	199.58104
3	25	mean	100.4	106.36	72.92	7.156	136.16	45.72	90.12	3.6264	7084
		sd	15.09967	14.982	10.9724	0.43458	20.9908	13.92413	19.17982	0.43472	953.90984
		se	3.01993	2.9965	2.19448	0.08691	4.19816	2.74482	3.83596	0.08694	190.78197
4	28	mean	92.78571	115.04	78.25	7.27857	128.143	42.67857	85.45429	3.79071	7144.64286
		sd	15.065	11.975	10.35659	0.39402	14.8413	13.40589	17.97713	0.52002	898.54438
		se	2.84701	2.263	1.95721	0.07446	2.80475	2.53347	3.39735	0.09827	169.80892

**Summary Table: 1B**

Class	Sample size	Statistic	N (%)	L (%)	E (%)	M (%)	Hb (gm%)	Tc (%)	Bc (%)
1	20	mean	55.45	32.85	9.65	2.05	11.82	65	23.1
		sd	4.225	3.454	3.825	0.805	0.621	4.278	4.3
		se	0.945	0.772	0.855	0.18	0.139	0.957	0.962
2	15	mean	53.667	36.4	5.667	4.267	10.647	63.733	23.6
		sd	5.907	4.514	3.32	1.289	1.447	5.234	5.414
		se	1.525	1.165	0.857	0.333	0.374	1.351	1.398
3	25	mean	588	31.56	7.64	2.64	10.704	62.16	25.32
		sd	6.203	5.375	4.906	1.353	1.144	7.22	5.732
		se	1.241	1.075	0.981	0.271	0.229	1.444	1.146
4	28	mean	55.536	33.571	8.857	2.107	11.064	63.5	23.929
		sd	5.635	4.346	3.71	0.673	1.257	7.287	4.705
		se	1.065	0.821	0.701	0.127	0.238	1.377	0.889

**Summary Table: 1C**

Class	Sample size	Statistic (mg%)	IgG (mg%)	IgA (mg%)	IgM (%)	Alb (%)	$\alpha_1$ (%)	$\alpha_2$ (%)	$\beta$ (%)	$\gamma$ (%)
1	20	mean	1519.3	199.65	114.35	73.985	3.095	3.135	4.83	14.885
		sd	199.158	19.132	28.079	5.513	1.28	1.042	1.36	4.62
		se	44.533	4.278	6.279	1.233	0.286	0.233	0.304	1.003
2	15	mean	2870.333	182.6	129.067	58.74	3.68	4.353	8.267	24.867
		sd	277.906	19.231	24.845	6.706	1.574	2.513	3.374	6.182
		se	71.755	4.965	6.415	1.732	0.407	0.649	0.871	1.596
3	25	mean	2777.96	190	135.84	67.844	3.792	3.74	5.084	19.48
		sd	410.31	15.773	33.308	3.86	1.406	1.484	1.606	3.417
		se	82.062	3.155	6.662	0.772	0.281	0.297	0.321	0.683
4	28	mean	2608.393	199.07	131.286	67.054	3.604	4.136	6.971	18.339
		sd	267.147	16.873	22.247	5.046	1.125	1.527	2.125	3.693
		se	50.486	3.189	4.204	0.954	0.213	0.289	0.402	0.698

**B: Analysis of Variance**

In this section, we will study whether or not the means of 21 blood parameters, BP (Systol/Diastol) and pulse rates are different for 4 classes of people described above. For this study, we consider the following two way analysis of a variance model using class and sex as two classifying factors for each of the 21 blood parameters, pulse rate and BP (Systol/Diastol):

$$(1) \quad y_{i,j,k}(t) = m + a(i) + b(j) + c(i,j) + e_{i,j,k}$$

where  $y_{i,j,k}(t)$  = response obtained from the  $k_{th}$  individuals for the  $j_{th}$  ( $j = 0, 1$ ) sex of the  $i_{th}$  ( $i = 1, 2, 3, 4$ ) class, corresponding to the parameter,  $t$ .

- $m$  = general effect
- $a(i)$  = effect due to  $i_{th}$  class ( $i = 1, \dots, 4$ )
- $b(j)$  = effect due to  $j_{th}$  sex ( $j = 0, 1$ )
- $a(i, j)$  = interaction effect between  $i_{th}$  class and  $j_{th}$  sex

Assumptions of the Analysis of variance model:

$$(a) \quad \sum_i a(i) = \sum_j b(j) = \sum_i c(i, j) = \sum_j c(i, j) = 0$$

(b)  $e(i, j, k)$  = independently and identically distributed as normal variate with mean zero and variance  $\sigma^2$ .

The assumptions of the model imply that  $y_{i,j,k(t)}$ 's are independently normally distributed with mean  $m + a(i) + b(j) + c(i, j)$  and variance  $\sigma^2$ . Since the number of observations of the 4 classes (20, 15, 28) are different, unbalanced analysis of variance techniques are used. This is clearly explained with an example by Kshirsagar (1983). Following the analysis of variance tables, the effects of  $\beta$ ,  $\gamma$ , IgG, Alb, Tp, Tch and Ech were found to be significant. This implies that the mean of each of the parameters  $\beta$ ,  $\gamma$ , IgG, Alb, Tp, Tch and Ech is different for 4 classes. The computations for these parameters are shown in the tables (2A -2G).

**Table 2A: ANOVA ( $\beta$ )**

Source	ss	df	ms	F-value	P-value
Class	123.106	3	41.035	8.141	0
Sex	0.303	1	0.303	0.06	0.807
Interaction	1.754	3	0.585	0.116	0.951
Error	403.26	80	5.041		

**Table 2B: ANOVA ( $\gamma$ )**

Source	ss	df	ms	F-value	P-value
Class	627.839	3	209.28	10.509	0
Sex	39.157	1	39.157	1.966	0.165
Interaction	35.099	3	11.7	0.588	0.625
Error	1593.121	80	19.914		

**Table 2C: ANOVA (IgG)**

Source	ss	df	ms	F value	P value
Class	14318260	3	4772753.3	47.721	0
Sex	68669.954	1	68669.954	0.687	0.41
Interaction	65112.199	3	3360328.2	33.599	0.844
Error	8001077.2	80	100013.47		



**Table 2D: ANOVA (Alb)**

Source	ss	df	ms	F value	P value
Class	1236.765	3	412.255	14.108	0
Sex	0.011	1	0.011	0	0.984
Interaction	31.364	3	10.455	0.358	0.784
Error	2337	80	29.222		

**Table 2E: ANOVA (TP)**

Source	ss	df	ms	F-value	P-value
Class	2.212	3	0.737	3.892	0.012
Sex	0.033	1	0.033	0.172	0.679
Interaction	0.465	3	0.155	0.817	0.488
Error	15.159	80	0.189		

**Table 2F: ANOVA (TCH)**

Source	SS	df	ms	F-value	P-value
Class	13386.909	3	4462.303	13.466	0
Sex	1.986	1	1.986	0.006	0.938
Interaction	1477.278	3	492.426	1.486	0.225
Error	26509.82	80	331.373		

**Table 2G: ANOVA (EC)**

Source	ss	df	ms	F-value	P-value
Class	14398.88	3	4799.627	15.454	0
Sex	272.46	1	272.46	0.877	0.352
Interaction	249.905	3	83.302	0.268	0.848
Error	24846.723	80	310.584		

### C. Correlation Matrix

The correlation matrix for 13 selected important parameters (viz. Alb, N,B,E, M, Hb, Tc, Tp, Tch, Ec), based on 88 observations obtained after combining 4 classes, are given in table 3. From table 3, we note that ‘*r*’ values of Ester Cholesterol (Ec) and Total Cholesterol (Tch), Bc and Alb are fairly high. But there is moderately high negative correlation between eosinophils (E) and neutrophils (N) and between neutrophils (N) and lymphocytes (L). Other correlations are quite low.

**Table 3: Correlation Matrix**

	Alb	N	L	E	M	Hb	Tc	Bc	Tp	Tch	Fc	Ec
Alb	1											
N	0.0244	1										
L	0.2996	-0.6653	1									
E	0.2996	-0.5307	-0.1506	1								
M	-0.452	0.06	0.1844	-0.2399	1							
Hb	0.2225	0.623	-0.1233	0.0466	-0.1339	1						
Tc	0.0945	0.571	-0.2196	0.0221	0.0337	0.1507	1					
Bc	0.0172	0.512	0.0443	-0.1387	0.0181	0.0709	0.0095	1				
Tp	0.1128	0.199	-0.005	0.2262	-0.1798	-0.0862	-0.053	-0.0983	1			
Tch	0.446	0.0666	-0.2467	0.2879	-0.4295	0.1672	0.087	-0.027	0.2972	1		
Fc	-0.125	0.0743	-0.0461	0.0439	-0.0533	-0.0817	-0.168	0.0101	0.0146	0.3247	1	
Ec	0.6435	0.1071	-0.2366	0.2906	-0.4317	0.2567	0.0279	-0.0307	0.2676	0.7868	-0.226	1

### D. Categorical Data Analysis

In this section we test whether or not, there is any association between (i) blood group and class and (ii) sex and class. Both the computed chi-squares for class vs sex were found to be insignificant (details are given in tables 4A and 4B). Thus neither the blood group nor sex has been found to be significantly associated with population class.

**Table 4A: Contingency table (Marginal subtable)**

Class	BG				Total
	o	a	b	ab	
Normal	4	7	1	8	20
Malarious	8	0	1	6	15
Febrile	8	7	3	7	25
Dormant	6	7	3	12	28
Total	26	21	8	33	88

Computed  $\chi^2 = 10.6316 < 16.919 = \chi^2 (0.05, 9)$ , BG = Blood Group

**Table 4B: Contingency table (Marginal subtable)**

Class	Sex		Total
	males	females	
Normal	3	17	20
Malarious	4	11	15
Febrile	10	15	25
Dormant	7	21	28
Total	24	64	88

Computed  $\chi^2 = 6.361 < 7.815 = \chi^2 (0.05, 3)$

### E. Predicting malaria infection

In this section, we try to find a method of classifying an individual in any of the 4 classes (Normal, Malarious, Febrile and Dormant) on the basis of the data collected on 23 blood parameters, Pulse rate, Blood pressure as well as sex and Blood group. From the categorical data analysis mentioned in section we note that Blood group and sex have no association with the class. In other words, sex and blood group cannot classify an individual in any of the four classes. But from the Analysis of variance technique given in section B, we note that Alb, Tp, Tch, Ech,  $\beta$ ,  $\gamma$ , IgG are different for different classes. We already mentioned in section C that Tch and Ech are highly correlated. So we use Tch along with  $\beta$ ,  $\gamma$ , IgG, Alb and Tp as classification variables. To classify an individual in any of the four classes, we compute  $L(x)$ , discriminant function or Fisher's classification function for each of the four classes (Seber, 1984) for some given value of  $x = (\beta, \gamma, \text{IgG}, \text{Alb}, \text{Tp}, \text{Tch})$ . So for each of the four classes, the discriminant function was computed by using the STATISTICA (discriminant analysis programme) assuming prior probabilities are the same for all the four classes (details are given in tables 5A and 5B). The discriminant functions are as follows:

$$\text{Class 1: } L_1(x) = -904.862 + 22.633\beta + 15.701\gamma + 0.006\text{IgG} + 16.115\text{Alb} + 44.378\text{Tp} - 0.386\text{Tch}$$

$$\text{Class 2: } L_2(x) = -928.602 + 23.594\beta + 16.262\gamma + 0.022\text{IgG} + 16.099\text{Alb} + 41.704\text{Tp} - 0.450\text{Tch}$$

$$\text{Class 3: } L_3(x) = -905.102 + 22.649\beta + 15.714\gamma + 0.021\text{IgG} + 15.867\text{Alb} + 41.172\text{Tp} - 0.321\text{Tch}$$

$$\text{Class 4: } L_4(x) = -922.172 + 23.180\beta + 15.820\gamma + 0.019\text{IgG} + 16.012\text{Alb} + 43.170\text{Tp} - 0.364\text{Tch}$$

**Table 5A**

STAT DISCRIM. ANALYSIS	Classification Functions		Grouping	CL (malaria status)
Variable	G1:1	G2:2	G3:3	G4:4
$\beta$	2.633	23.594	22.649	23.180
$\gamma$	5.701	16.262	15.714	15.820
IgG	0.006	0.002	0.021	0.019
Alb	16.115	16.099	15.867	16.012
Tp	44.378	41.704	41.172	43.170
Tch	-0.386	-0.450	-0.321	-0.364
Constant	-904.862	-928.602	-905.102	-922.172

**Table 5B**

STAT DISCRIM. ANALYSIS	Classification Matrix (malaria. sta)				
	Percent	Rows observed classifications Columns: Predicted classifications			
Group	Correct	G1:1 p=0.22727	G2:2 p=0.17045	G3:3 p=0.28409	G4:4 p=0.31818
G1:1	100.0000	20	0	0	0
G2:2	86.6667	0	13	1	1
G3:3	76.0000	0	0	19	6
G4:4	75.0000	0	1	6	21
Total	82.9545	20	14	26	28

The method involves the assignment of assigning an individual with  $x = (\beta, \gamma, \text{IgG}, \text{Alb}, \text{Tp}, \text{Tch})$  to the class with the longest value of the discriminant function. For example, when  $\beta = 1, \gamma = 8, \text{IgG} = 2000, \text{Alb} = 60, \text{Tp} = 6.5$  and  $\text{Tch} = 130$ , we have  $x = (1, 8, 2000, 60, 6.5, 130)$  and  $L_1(x) = 460.556, L_2(x) = 447.604, L_3(x) = 463.269, L_4(x) = 459.573$ . So, an individual having  $\beta = 1, \gamma = 8, \text{IgG} = 2000, \text{Alb} = 60, \text{Tp} = 6.5$  and  $\text{Tch} = 130$  will be classified in class 3 (Febrile) since  $L_3(x) = 463.269$  and is the maximum. In the process of classification, there is the possibility of misclassification. The estimated percentages of correct classification were computed by STATISTICA and it yielded 82.9545% for the total and 100%, 86.6%, 76% and 75% respectively for class 1, class 2, class 3 and class 4 respectively.

**Remark 1:** Seshadri *et al.* (1983), Sharma *et al.* (1983) indicated the significant role of Albumin and Total Cholesterol in detecting malaria infected cases.

□

## F. Multiple regression

In this section, we have presented multiple regression for some of blood parameters discussed above.

Dependent Variable	Independent Variable	Multiple regression
WBC	N,L,E,M	WBC=10594.4– 35.69N –44.03L +46.33E– 120.83M
RBC	Hb, Tc, Bc, IgG, IgM, IgA	RBC=0.13+0.365 Hb* –0.0006 Tc +0.002 Bc +0.00002 IgG –0.00009 IgM –0.0009 IgA
Tp	Alb, $\alpha_1$ , $\alpha_2$ , $\beta$ , $\gamma$	Tp =2.303+0.051 Alb* +0.035 $\alpha_1$ +0.045 $\alpha_2$ +0.046 $\beta$ =0.042 $\gamma$ *
Tch	Fc, Ec	Tch=14.260+0.874Fc* +0.8823 Ec*

\* indicates significant (at 5% level) regression coefficient corresponding to the respective independent variable.

From the above table, we note that none of the independent variables N, L, B, M can be used effectively for predicting size of WBC since regression coefficients corresponding to the variables are independent. In a similiar fashion we we can conclude that only Hb, Alb,  $\gamma$  and both Ec and Bc can be used as reliable predictor for RBC, Tp, and Tch respectively.

## 4. CONCLUSION

The survey based on cross-sectional data analysis on the tribal population described in this paper revealed that the 5 blood parameters viz. Tp, Alb, Tch, IgG,  $\beta$  and  $\gamma$ , out of the 21 blood parameters considered, may be used for predicting malarial endemicity of the tribal community. The factors like blood group and sex do not take important roles for the prediction of malaria. There are fairly high positive correlations between Tch and Ec: Alb and Ec. But there are also moderately high negative correlations between R and N and N and M. The study is based on limited data. More accurate model making require the collection of further data, especially of females who are recalcitrant to respond due to their «social taboos».

## ACKNOWLEDGEMENTS

We would like to thank the referee for his useful suggestions that greatly improved this work. We also thank Mr. A. Bose, Mr. P.P. Chakravorty and Mr. C. Kundu for statistical computations. Financial support was obtained from the Indian Council of Medical Research and the Indian Statistical Institute, Calcutta for conducting the survey.

## REFERENCES

- [1] **Blood, B. Ed.** (1977). *In vitro methods in cell-mediated immunity*. Academic Press, New York.
- [2] **Davie, J.V. and Lewis, S.M.** (1984). *Practical Haematology*. Churchill Livingstone.
- [3] **Kshirsagar, A.M.** (1983). *A course in Linear Models*. Marcell Dekkar Inc., New York, 305–307.
- [4] **Lowry, O.H., Rosenbrough, N.J., Farr, A.L. and Randall, R.J.** (1951). «Protein measurement with Folin Phenol Reagent». *Journal of Biological Chemistry*, **193**, 265–275.
- [5] **Mancini, G., Carbonara, A.O. and Horemans, J.F.** (1965). «Immunochemical quantitation of antigens by single radial immuno-diffusion». *Immunochemistry*, **2**, 235.
- [6] **Seber, G.A.F.** (1984). *Multivariate Observations*. John Wiley & Sons, USA 332–335.
- [7] **Sheshadri, C., Shetty B. Ramkrishna, Gouri, N., Sitaraman, R., Revathi, R., Venkatraghaban, S. and Chari, M.V.** (1983). «Biochemical changes at different levels of parasitaemia in *Plasmodium vivax* Malaria». *Indian Journal of Medical Research*, **77**, 437–442.
- [8] **Sharma, S., Shrivastava, I.K., Amarnath Dutta, G.P. and Agarwal S.S.** (1983). «Serum proteins and Immunoglobulin changes in Human Malaria». *Indian Journal of Malariology*, **20**, 15–19.
- [9] **Smithies, O.** (1955). «Zone electrophoresis in starch gels: Group variations in the serum protein of normal human adults». *Biochemical Journal*, **61**, 629.
- [10] **Wootton, I.D.P.** (1974). *Micro-analysis in Medical Biochemistry*. 4th ed. (J. and A. Churchill Ltd.) 83–86.